Journal of General Microbiology (1989), **135**, 1521–1528. *Printed in Great Britain*

Nucleotide Sequence of the Neopullulanase Gene from *Bacillus stearothermophilus*

By TAKASHI KURIKI AND TADAYUKI IMANAKA*

Department of Fermentation Technology, Faculty of Engineering, Osaka University, Yamadaoka, Suita-shi, Osaka 565, Japan

(*Received 15 December 1988; revised 1 March 1989; accepted 20 March 1989*)

The gene (*nplT*) for a new type of pullulan-hydrolysing enzyme, neopullulanase, from *Bacillus stearothermophilus* TRS40 was sequenced. The DNA sequence revealed only one large open reading frame, composed of 1764 bases and 588 amino acid residues (M_r 69144). Although the thermostable neopullulanase contained eight cysteine residues, they did not provide conformational stability by disulphide bonds. A comparison was made of the amino acid sequences of α-amylase, neopullulanase, isoamylase, pullulanase and cyclodextrin glucanotransferase. All the enzymes examined contained four highly conserved regions which probably constitute the active centres of the enzymes. The amino acid residues required for the specificity of neopullulanase are compared with those of α-amylase and other amylolytic enzymes.

INTRODUCTION

Four types of pullulan-hydrolysing enzymes have been described (Imanaka & Kuriki, 1989): (i) glucoamylase (glucan 1,4-α-glucosidase; EC 3.2.1.3), which hydrolyses pullulan from nonreducing ends to produce glucose; (ii) pullulanase (α-dextrin endo-1,6-α-glucosidase; EC 3.2.1.41), which hydrolyses α-(1→6)-glucosidic linkages of pullulan to produce maltotriose; (iii) isopullulanase (EC 3.2.1.57), which hydrolyses α-(1→4)-glucosidic linkages of pullulan to produce isopanose (6-*O*-α-maltosylglucose); and (iv) neopullulanase, which hydrolyses α-(1→4)-glucosidic linkages of pullulan to produce panose (6^2-*O*-α-glucosylmaltose). The last enzyme was reported to be a new type of pullulan-hydrolysing enzyme from *Bacillus stearothermophilus* TRS40 (Kuriki *et al.*, 1988*a*). We have cloned the genes for the neopullulanase and pullulanase in *Bacillus subtilis* (Kuriki *et al.*, 1988*a*, *b*, respectively). The neopullulanase could hydrolyse pullulan efficiently and only hydrolysed a small amount of starch. Further studies revealed that neopullulanase hydrolysed not only α-(1→4)-glucosidic linkages but also specific α-(1→6)-glucosidic linkages of several branched oligosaccharides (Imanaka & Kuriki, 1989).

It was previously demonstrated that four highly conserved regions existed in eleven different α-amylases (EC 3.2.1.1) (Nakajima *et al.*, 1986) and the regions were most likely to be the active and/or substrate-binding sites of the enzymes (Matsuura *et al.*, 1984). These homologous regions have also been discussed for other kinds of amylolytic enzymes (Amemura *et al.*, 1988; Binder *et al.*, 1986; Kimura *et al.*, 1987; McPherson & Charalambous, 1988; Sakai *et al.*, 1987).

This paper describes the nucleotide sequence of the cloned neopullulanase gene. To further investigate the unique action pattern of the neopullulanase at the molecular level, the amino acid sequence of the neopullulanase is compared with those of α-amylase, isoamylase, pullulanase and cyclodextrin glucanotransferase, and the specificity of the neopullulanase is described.

METHODS

Bacterial strains, plasmids and phages. The *Bacillus* strain used was *Bacillus subtilis* NA-1 (*arg-15 hsmM hsrM* Amy^- Npr^-) (Kuriki *et al.*, 1988*a*). *Escherichia coli* MV1184 [*ara* Δ(*lac–pro*) *strA thi* (φ80Δ*lacIZ*ΔM15) Δ(*srl–recA*)306::Tn*10*(Tet^r); F′:*traD36 proAB lacI*^{q}ZΔM15] was used as a host for plasmids pUC118 and

pUC119 and for phage M13KO7 (Vieira & Messing, 1987). Plasmid pPP10 [Tc^r *nplT*$^+$ (structural gene of the neopullulanase from *B. stearothermophilus* TRS40)] was described previously (Kuriki *et al.*, 1988*a*).

Plasmid isolation, restriction enzyme treatment, and ligation of DNA. Plasmid DNA was prepared by either the rapid alkaline extraction method or CsCl/ethidium bromide equilibrium density gradient centrifugation as described previously (Imanaka *et al.*, 1982). Treatment of DNA with restriction enzymes and ligation of DNA with T4 DNA ligase were done as recommended by the manufacturer.

Transformation. Transformation of *E. coli* with plasmid DNA, and transformation of competent *B. subtilis* cells, were done as described previously (Imanaka *et al.*, 1981).

Gel electrophoresis for DNA analysis and isolation. For the analysis of DNA, gel electrophoresis with agarose or polyacrylamide was done under standard conditions (Maniatis *et al.*, 1982). A GENECLEAN kit (BIO 101, La Jolla, Calif., USA) was used for the recovery of DNA from agarose.

DNA sequencing analysis. DNA sequencing was done by the dideoxy chain-terminating method (Sanger *et al.*, 1977) with a SEQUENASE kit (United State Biochemical Corporation, Cleveland, Oh., USA). For preparing single-stranded DNA, pUC118/119 and helper phage M13KO7 were used (Vieira & Messing, 1987). Some strands were analysed by the dideoxy sequencing method using alkaline denatured plasmid templates (Hattori & Sakaki, 1986). Although some parts of the DNA were analysed from one strand, those sequences could be determined without any ambiguity on the sequencing gels, because the substitution of dGTP by the nucleotide analogue dITP with the SEQUENASE kit was successful in elimination of compressed regions. All restriction sites were overlapped by sequencing across them.

Detection of neopullulanase-producing colonies and assay of neopullulanase activity. Neopullulanase-producing colonies were selected on PLL agar plates and neopullulanase activity was assayed at 50 °C as described previously (Kuriki *et al.*, 1988*a*).

Purification of neopullulanase. B. subtilis NA-1(pPP10) was cultivated in LSII broth (Kuriki *et al.*, 1988*a*) at 37 °C for 16 h. The culture supernatant was used for enzyme purification. Purification procedures for the neopullulanase were described previously (Kuriki *et al.*, 1988*a*).

Amino acid composition and amino-terminal amino acid sequence. Using neopullulanase purified to homogeneity, the amino acid composition was analysed as described previously (Matsumura *et al.*, 1984). The amino-terminal amino acid sequence was analysed by Edman degradation as described elsewhere (Takagi *et al.*, 1985).

Enzymes and chemicals. Restriction endonucleases, T4 DNA ligase, and DNA polymerase I (Klenow fragment) were purchased from Takara Shuzo Co., Kyoto, Japan. Tetracycline was from Sigma. All other chemicals used were from Wako Pure Chemical Industries, Osaka, Japan.

Analysis of amino acid sequence homology. This was done with an NEC PC-9801 computer (Nippon Electric Co.) and the GENETYX system (Software Development Co., Tokyo, Japan).

RESULTS

Nucleotide sequence of the neopullulanase gene

We have previously shown that the structural gene for the neopullulanase (*nplT*) from *B. stearothermophilus* TRS40 is cloned in a 3·4 kb (2·2 MDa) *Hin*dIII fragment of plasmid pPP10 (Kuriki *et al.*, 1988*a*). Fig. 1 shows the restriction map of the 3·4 kb *Hin*dIII fragment. Using a frame-shift mutation generated by end-filling with DNA polymerase I, we confirmed that a single *Sal*I site was located in *nplT*. The nucleotide sequence of *nplT* was determined (Fig. 2), according to the strategy shown in Fig. 1.

To correlate the nucleotide sequence data with the structure of the neopullulanase, the amino-terminal amino acid sequence of the enzyme from *B. subtilis* NA-1(pPP10) was determined through five cycles of the Edman degradation procedure (Allen, 1981). The first five amino acids of the enzyme were Met–Arg–Lys–Glu–Ala. This sequence completely matched that deduced from the nucleotide sequence only when ATG was taken as the initiation codon at position +1 (Fig. 2).

Starting from the ATG codon at position +1 and terminating in a TAG nonsense codon at position +1765, the single open reading frame was composed of 1764 nucleotides (Fig. 2). The maximum length of other open reading frames was 474 nucleotides in the two other reading frames of the same strand and the three reading frames of the complementary strand.

At 6 bases upstream from the ATG codon, there was a 10-base sequence, AAGGAGGAGA (−16 to −7), which exhibited complementarity with the 3′ end of the 16S rRNA from *B. stearothermophilus* (Kozak, 1983) and *B. subtilis* (Moran *et al.*, 1982); hence it is the most probable ribosome-binding site (Shine–Dalgarno sequence) of *nplT*. The free energy of

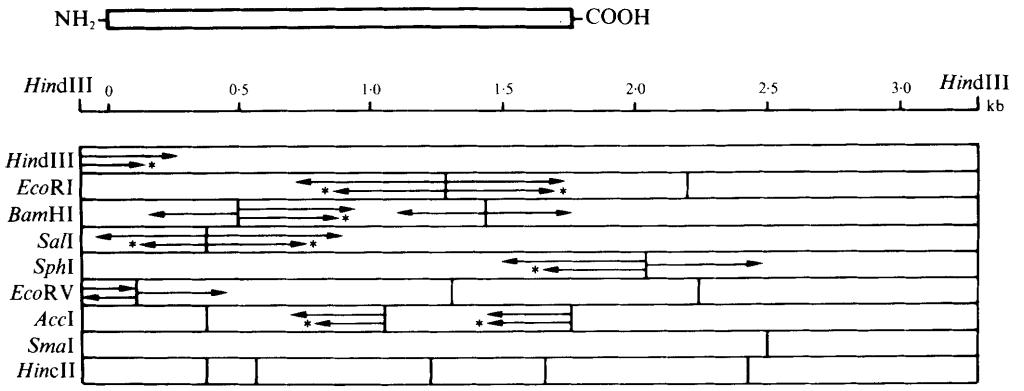


Fig. 1. Restriction map and sequencing strategy of the 3·4 kb *Hin*dIII fragment of pPP10. The location and size of the neopullulanase gene are shown by the open bar. Arrows indicate the direction and length of individual sequence determinations. Asterisks indicate the regions analysed using dITP.

formation of the most stable pairing was calculated as $-18{\cdot}0$ kcal mol^{-1} ($-75{\cdot}3$ kJ mol^{-1}) (Tinoco *et al.*, 1973).

Two putative promoters (-35 and -10 regions) are shown in Fig. 2; one is TTGACTTTTTTCTCTCTTTTCGTACTCT (-90 to -62) and the other TTTTTTCTCTCTTTTCGTACTCTTTAAT (-85 to -57). However, these two putative promoters are somewhat different from the consensus sequence for the σ^{43} factor (TTGACA for the -35 region and TATAAT for the -10 region) in *B. subtilis* (Moran *et al.*, 1982). The distance between the -35 region and the -10 region was 17 bp for each promoter, which is the same as the consensus distance in *B. subtilis* (Moran *et al.*, 1982). Furthermore, 39 bases downstream from the termination codon at position $+1765$ was a palindromic sequence (Fig. 2) that might act as a transcription terminator (Rosenberg & Court, 1979).

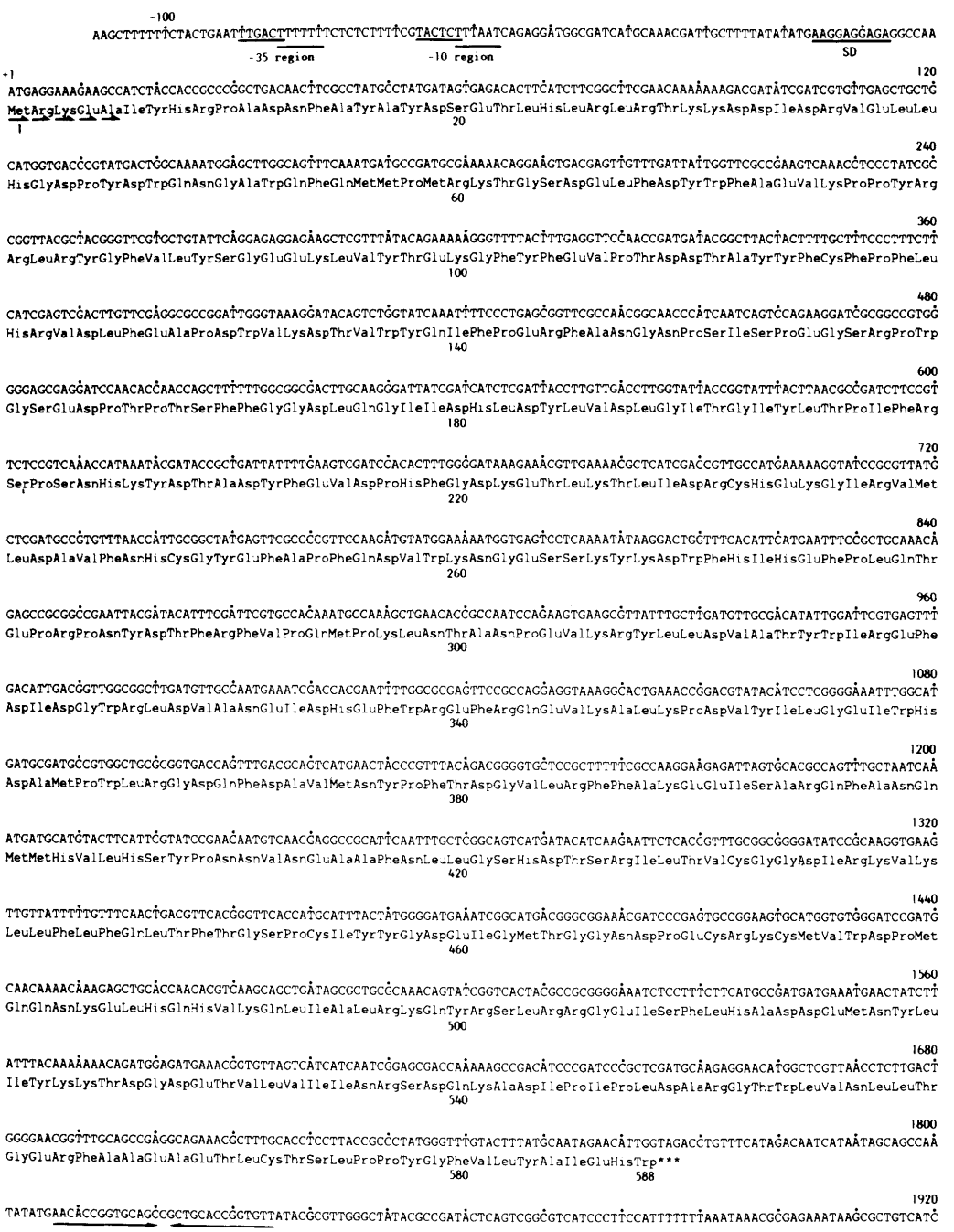
Amino acid analysis

The M_r of the neopullulanase, calculated as 69144 from the amino acid sequence (588 amino acid residues), is in reasonable agreement with the direct assessment (M_r 62000) by SDS polyacrylamide gel electrophoresis (Kuriki *et al.*, 1988*a*). The amino acid composition of the neopullulanase agreed with that deduced from the nucleotide sequence (data not shown). The codon usage was rather randomly distributed, without clear bias for any particular third-position base, and was nearly the same as that of *B. stearothermophilus* α-amylase (Nakajima *et al.*, 1985).

The neopullulanase contained eight cysteine residues (Fig. 2). If some disulphide bonds existed, they might provide conformational stability (Perry & Wetzel, 1984). To examine whether disulphide bonds were present in the neopullulanase, we compared the thermostability of the enzyme in the presence or absence of a reducing agent. 2-Mercaptoethanol (final concentration 10 mM) was added to the enzyme solution and the thermostability at 60, 65 and 70 °C was measured as described previously (Kuriki *et al.*, 1988*a*). The thermostability and specific activity of the reduced neopullulanase were rather higher than those of the native (non-reduced) enzyme, although the reason is not clear yet (data not shown). This result implied that the neopullulanase did not contain any disulphide bonds necessary for enzyme activity or for stability.

Comparison of the amino acid sequences of homologous regions in several amylolytic enzymes

We analysed the homologous amino acid sequences for α-amylases (Nakajima *et al.*, 1986), the neopullulanase, isoamylase (Amemura *et al.*, 1988), pullulanase (Katsuragi *et al.*, 1987) and cyclodextrin glucanotransferases (Binder *et al.*, 1986; Kimura *et al.*, 1987; Sakai *et al.*, 1987). As


-100
AAGCTTTTTTCTACTGAATTTGACTTTTTTTCTCTCTTTTCGTACTCTTTAATCAGAGGATGGCGATCATGCAAACGATTGCTTTTATATATGAAGGAGGAGAGGCCAA
-35 region
-10 region
SD
+1
ATGAGGAAAGAAGCCATCTACCACCGCCCGGCTGACAACTTCGCCTATGCCTATGATAGTGAGACACTTCATCTTCGGCTTCGAACAAAAAAAGACGATATCGATCGTGTTGAGCTGCTG
MetArgLysGluAlaIleTyrHisArgProAlaAspAsnPheAlaTyrAlaTyrAspSerGluThrLeuHisLeuArgLeuArgThrLysLysAspAspIleAspArgValGluLeuLeu
1
20
CATGGTGACCCGTATGACTGGCAAAATGGAGCTTGGCAGTTTCAAATGATGCCGATGCGAAAAACAGGAAGTGACGAGTTGTTTGATTATTGGTTCGCCGAAGTCAAACCTCCCTATCGC
HisGlyAspProTyrAspTrpGlnAsnGlyAlaTrpGlnPheGlnMetMetProMetArgLysThrGlySerAspGluLeuPheAspTyrTrpPheAlaGluValLysProProTyrArg
60
CGGTTACGCTACGGGTTCGTGCTGTATTCAGGAGAGGAGAAGCTCGTTTATACAGAAAAAGGGTTTTACTTGAGGTTCCAACCGATGATACGGCTTACTACTTTTGCTTTCCCTTTCTT
ArgLeuArgTyrGlyPheValLeuTyrSerGlyGluGluLysLeuValTyrThrGluLysGlyPheTyrPheGluValProThrAspAspThrAlaTyrTyrPheCysPheProPheLeu
100
CATCGAGTCGACTTGTTCGAGGCGCCGGATTGGGTAAAGGATACAGTCTGGTATCAAATTTTCCCTGAGCGGTTCGCCAACGGCAACCCATCAATCAGTCCAGAAGGATCGCGGCCGTGG
HisArgValAspLeuPheGluAlaProAspTrpValLysAspThrValTrpTyrGlnIlePheProGluArgPheAlaAsnGlyAsnProSerIleSerProGluGlySerArgProTrp
140
GGGAGCGAGGATCCAACACCAACCAGCTTTTTTGGCGGCGACTTGCAAGGGATTATCGATCATCTCGATTACCTTGTTGACCTTGGTATTACCGGTATTTACTTAACGCCGATCTTCCGT
GlySerGluAspProThrProThrSerPhePheGlyGlyAspLeuGlnGlyIleIleAspHisLeuAspTyrLeuValAspLeuGlyIleThrGlyIleTyrLeuThrProIlePheArg
180
TCTCCGTCAAACCATAAATACGATACCGCTGATTATTTTGAAGTCGATCCACACTTTGGGGATAAAGAAACGTTGAAAACGCTCATCGACCGTTGCCATGAAAAAGGTATCCGCGTTATG
SerProSerAsnHisLysTyrAspThrAlaAspTyrPheGluValAspProHisPheGlyAspLysGluThrLeuLysThrLeuIleAspArgCysHisGluLysGlyIleArgValMet
220
CTCGATGCCGTGTTTAACCATTGCGGCTATGAGTTCGCCCCGTTCCAAGATGTATGGAAAAATGGTGAGTCCTCAAAATATAAGGACTGGTTTCACATTCATGAATTTCCGCTGCAAACA
LeuAspAlaValPheAsnHisCysGlyTyrGluPheAlaProPheGlnAspValTrpLysAsnGlyGluSerSerLysTyrLysAspTrpPheHisIleHisGluPheProLeuGlnThr
260
GAGCCGCGGCCGAATTACGATACATTTCGATTCGTGCCACAAATGCCAAAGCTGAACACCGCCAATCCAGAAGTGAAGCGTTATTTGCTTGATGTTGCGACATATTGGATTCGTGAGTTT
GluProArgProAsnTyrAspThrPheArgPheValProGlnMetProLysLeuAsnThrAlaAsnProGluValLysArgTyrLeuLeuAspValAlaThrTyrTrpIleArgGluPhe
300
GACATTGACGGTTGGCGGCTTGATGTTGCCAATGAAATCGACCACGAATTTTGGCGCGAGTTCCGCCAGGAGGTAAAGGCACTGAAACCGGACGTATACATCCTCGGGGAAATTTGGCAT
AspIleAspGlyTrpArgLeuAspValAlaAsnGluIleAspHisGluPheTrpArgGluPheArgGlnGluValLysAlaLeuLysProAspValTyrIleLeuGlyGluIleTrpHis
340
GATGCGATGCCGTGGCTGCGCGGTGACCAGTTTGACGCAGTCATGAACTACCCGTTTACAGACGGGGTGCTCCGCTTTTTCGCCAAGGAAGAGATTAGTGCACGCCAGTTTGCTAATCAA
AspAlaMetProTrpLeuArgGlyAspGlnPheAspAlaValMetAsnTyrProPheThrAspGlyValLeuArgPhePheAlaLysGluGluIleSerAlaArgGlnPheAlaAsnGln
380
ATGATGCATGTACTTCATTCGTATCCGAACAATGTCAACGAGGCCGCATTCAATTTGCTCGGCAGTCATGATACATCAAGAATTCTCACCGTTTGCGGCGGGGATATCCGCAAGGTGAAG
MetMetHisValLeuHisSerTyrProAsnAsnValAsnGluAlaAlaPheAsnLeuLeuGlySerHisAspThrSerArgIleLeuThrValCysGlyGlyAspIleArgLysValLys
420
TTGTTATTTTTGTTTCAACTGACGTTCACGGGTTCACCATGCATTTACTATGGGGATGAAATCGGCATGACGGGCGGAAACGATCCCGAGTGCCGGAAGTGCATGGTGTGGGATCCGATG
LeuLeuPheLeuPheGlnLeuThrPheThrGlySerProCysIleTyrTyrGlyAspGluIleGlyMetThrGlyGlyAsnAspProGluCysArgLysCysMetValTrpAspProMet
460
CAACAAAACAAAGAGCTGCACCAACACGTCAAGCAGCTGATAGCGCTGCGCAAACAGTATCGGTCACTACGCCGCGGGGAAATCTCCTTTCTTCATGCCGATGATGAAATGAACTATCTT
GlnGlnAsnLysGluLeuHisGlnHisValLysGlnLeuIleAlaLeuArgLysGlnTyrArgSerLeuArgArgGlyGluIleSerPheLeuHisAlaAspAspGluMetAsnTyrLeu
500
ATTTACAAAAAAACAGATGGAGATGAAACGGTGTTAGTCATCATCAATCGGAGCGACCAAAAAGCCGACATCCCGATCCCGCTCGATGCAAGAGGAACATGGCTCGTTAACCTCTTGACT
IleTyrLysLysThrAspGlyAspGluThrValLeuValIleIleAsnArgSerAspGlnLysAlaAspIleProIleProLeuAspAlaArgGlyThrTrpLeuValAsnLeuLeuThr
540
GGGGAACGGTTTGCAGCCGAGGCAGAAACGCTTTGCACCTCCTTACCGCCCTATGGGTTTGTACTTTATGCAATAGAACATTGGTAGACCTGTTTCATAGACAATCATAATAGCAGCCAA
GlyGluArgPheAlaAlaGluAlaGluThrLeuCysThrSerLeuProProTyrGlyPheValLeuTyrAlaIleGluHisTrp***
580
588
TATATGAACACCGGTGCAGCCGCTGCACCGGTGTTATACGCGTTGGGCTATACGCCGATACTCAGTCGGCGTCATCCCTTCCATTTTTTTAAATAAACGCGAGAAATAAGCGCTGTCATC
AATGCCAACTCGTTTGGCGATGGCCTCGACCGTTTCATTTGTTTCAGCCAACCATTGTTTCGCCTTTGACAATCGGTAGTACGCCAAATAATGGCTAAAGCCCATCCCGAGCGTTCTTTG
CATGC

		Region 1		Region 2		Region 3		Region 4	
			D A V I N H		G F R L D A A K H		E V I D		F V D N H D
AMY	Consensus								
AMY	*A. ory.*	117	D V V A N H	202	G L R I D T V K H	230	E V L D	292	F V E N H D
AMY	*B. stearo.*	101	D V V F D H	230	G F R L D A V K H	264	E Y W S	326	F V D N H D
AMY	*B. amylo.*	98	D V V L N H	227	G F R I D A A K H	261	E Y W Q	323	F V E N H D
AMY	*B. sub.*	97	D A V I N H	172	G F R F D A A K H	208	E I L Q	264	W V E S H D
AMY	Rat	96	D A V I N H	190	G F R L D A A K H	230	E V I D	292	F V D N H D
AMY	Mouse, s	96	D A V I N H	193	G F R L D A S K H	233	E V I D	295	F V D N H D
AMY	Mouse, p	96	D A V I N H	190	G F R L D A A K H	230	E V I D	292	F V D N H D
AMY	Hog	96	D A V I N H	193	G F R I D A S K H	233	E V I D	295	F V D N H D
AMY	Human, s	99	D A V I N H	196	G F R I D A S K H	236	E V I D	298	F V D N H D
AMY	Human, p	99	D A V I N H	196	G F R L D A S K H	236	E V I D	298	F V D N H D
AMY	Barley	101	D I V I N H	127	D G R L D W G P H	218	E V W D	299	F V D N H D
NPL	*B. stearo.*	242	D A V F N H	324	G W R L D V A N E	357	E I W H	419	L L G S H D
IAM	*P. amylo.*	291	D V V Y N H	370	G F R F D L A S V	454	E W S V	502	F I D V H D
PUL	*K. aero.*	600	D V V Y N H	671	G F R F D L M G Y	704	E G W D	827	Y V S K H D
CGT	*K. pne.*	130	D Y A D N H	219	A I R I D A I K H	257	E W F G	328	F M D N H D
CGT	*B. mace.*	135	D F A P N H	225	G I R F D A V K H	258	E W F L	324	F I D N H D
CGT	Alk. *B.*	135	D F A P N H	225	G I R V D A V K H	268	E Y H Q	323	F I D N H D
CGT	*B. stearo.*	131	D F A P N H	221	G I R M D A V K H	253	E W F L	319	F I D N H D

Fig. 3. Highly conserved regions among α-amylase, neopullulanase, isoamylase, pullulanase and cyclodextrin glucanotransferase. The amino acid residues identical with the consensus sequence of α-amylase are shaded. Amino acid residues are shown by single letters as follows: A, Ala; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr. Enzymes are abbreviated as: AMY, α-amylase; NPL, neopullulanase; IAM, isoamylase; PUL, pullulanase; CGT, cyclodextrin glucanotransferase. Enzyme sources are abbreviated as: *A. ory.*, *Aspergillus oryzae*; *B. stearo.*, *Bacillus stearothermophilus*; *B. amylo.*, *Bacillus amyloliquefaciens*; *B. sub.*, *Bacillus subtilis*; *P. amylo.*, *Pseudomonas amyloderamosa*; *K. aero.*, *Klebsiella aerogenes*; *K. pne.*, *Klebsiella pneumoniae*; *B. mace.*, *Bacillus macerans*; Alk. *B.*, alkalophilic *Bacillus*; s, saliva; p, pancreas. For the reference of each amino acid sequence, see the text. Numbering of the amino acid sequences of the enzymes starts at the amino-terminal amino acid of each mature enzyme. Active sites and substrate binding sites proposed by Matsuura *et al.* (1984) for Taka-amylase A from *A. oryzae* are indicated by ○ and □, respectively.

the results demonstrate (Fig. 3), the four highly conserved regions previously demonstrated in α-amylases were also found in the neopullulanase, isoamylase, pullulanase and cyclodextrin glucanotransferases. The arrangement of regions 1, 2, 3 and 4 from the amino-termini among these enzymes were the same. The other parts of the neopullulanase showed almost no similarity to those of α-amylases, isoamylase, pullulanase and cyclodextrin glucanotransferases.

DISCUSSION

We have determined the complete nucleotide sequence of the gene for a new type of pullulan-hydrolysing enzyme, designated neopullulanase, from *B. stearothermophilus* TRS40. An open reading frame, composed of 1764 bases and 588 amino acid residues, was found. The thermostable neopullulanase contained eight cysteine residues. However, they did not provide conformational stability through disulphide bond formation.

Fig. 2. Nucleotide and amino acid sequences of the neopullulanase gene. The nucleotide sequence is presented from the *Hin*dIII site (nucleotide −109) to the *Sph*I site (nucleotide +2045). The nucleotide sequence is counted from the first base of open reading frame. The amino acid sequence is shown beneath the nucleotide sequence. The amino-terminal amino acid sequence of the neopullulanase, determined by the Edman method, is denoted by half-head arrows. The first amino acid of the translation (Met) is counted as 1. A probable Shine–Dalgarno sequence (nucleotide −16 to −7) and putative promoter regions (−35 and −10 regions) are shown by solid lines below the nucleotide sequence. Asterisks indicate a stop codon. The sequence containing inverted repeat structures downstream from the termination codon TAG is designated by →←.

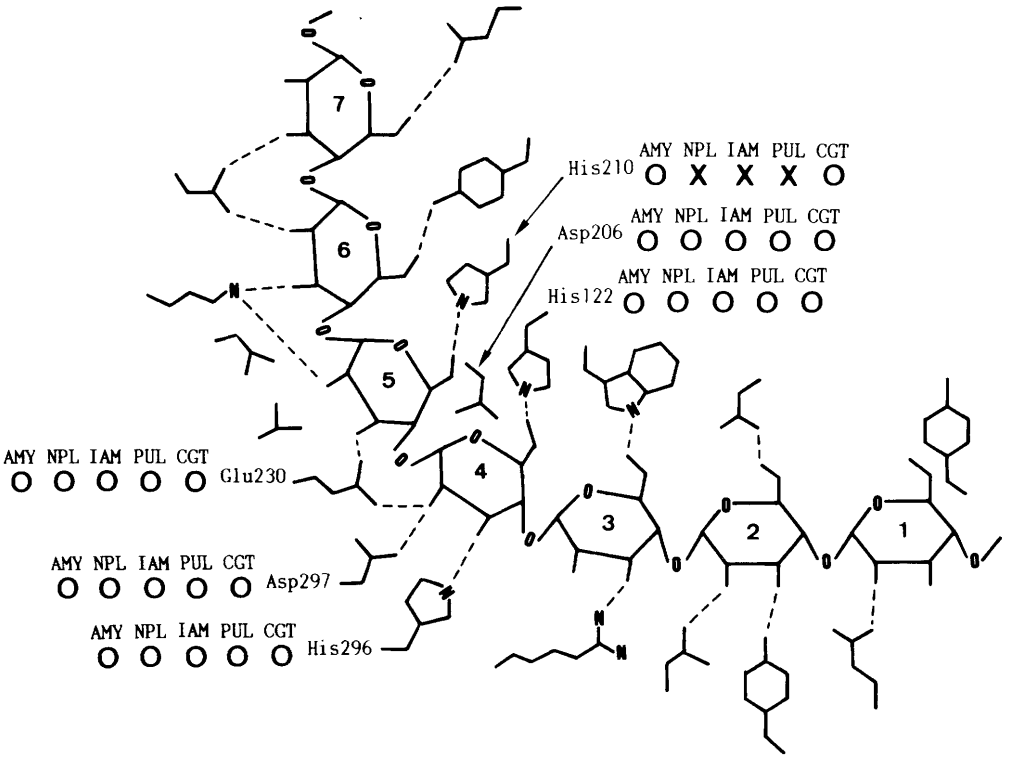


Fig. 4. Proposed substrate-binding model of Taka-amylase A (Matsuura *et al.*, 1984) and conservation of corresponding amino acid residues in other α-amylases, neopullulanase, isoamylase, pullulanase and cyclodextrin glucanotransferases (see Fig. 3). The glucose residues of amylose (as a substrate) are specified by the numbers 1 to 7. Enzymes are abbreviated as in Fig. 3. ○, Conserved residue; ×, non-conserved residue.

Table 1. *Action of various amylolytic enzymes*

Relative hydrolysis and transglucosylation rates are indicated as follows: + +, strong; +, medium; ±, weak; −, none. α-(1→4), α-(1→4)-glucosidic linkage; α-(1→6), α-(1→6)-glucosidic linkage.

	Pullulan		Starch/amylopectin		Glycogen		
Enzyme	α-(1→4)	α-(1→6)	α-(1→4)	α-(1→6)	α-(1→4)	α-(1→6)	Trans-glucosylation
α-Amylase	−	−	+ +	−			±
Neopullulanase	+ +	+	±	±			
Isoamylase	−	−	−	+ +	−	+ +	
Pullulanase	−	+ +	−	+ +	−	±	
Cyclodextrin glucanotransferase	−	−	+ +	−			+ +

Although most of the amylolytic enzymes are secretory proteins, no typical signal sequence was found in the amino-terminal region of the neopullulanase. When *B. subtilis* NA-1(pPP10) was cultivated for 13 h (early stationary phase), about 20% of the total neopullulanase activity was detected in the culture supernatant. Since a major portion of the activity (about 80%) was found in the cell extract, which was prepared by sonication (19·5 KHz, 10 min), we purified the intracellular neopullulanase, and the amino-terminal amino acid sequence was compared with that of the enzyme purified from the culture supernatant. The sequences of the first five amino

acids of the enzymes from the two different preparations were identical. Since the neopullulanase was exclusively extracellular in *B. stearothermophilus* TRS40, the different localization of the neopullulanase in *B. subtilis* remains to be investigated.

We found four highly conserved regions in α-amylase, cyclodextrin glucanotransferase, and even in the neopullulanase, isoamylase and pullulanase, whose different patterns of action are summarized in Table 1. This finding might be very interesting to point out the homologies between α-(1→4)-glucanohydrolases (α-amylase and cyclodextrin glucanotransferase) and α-(1→6)-glucanohydrolases (isoamylase and pullulanase). Pullulanase hydrolyses pullulan, whereas isoamylase does not. The neopullulanase hydrolyses not only α-(1→4)- but also α-(1→6)-glucosidic linkages (Imanaka & Kuriki, 1989; Kuriki *et al.*, 1988*a*). The enzyme could hydrolyse pullulan efficiently and only a small amount of starch (Kuriki *et al.*, 1988*a*). Therefore, the neopullulanase is a novel enzyme which might be classified between α-amylase and pullulanase from the viewpoint of its action pattern.

In this context, we analysed the homologies of other glucanohydrolases, such as glucoamylase (EC 3.2.1.3) (Yamashita *et al.*, 1985), β-amylase (EC 3.2.1.2) (Rhodes *et al.*, 1987), cellulase (EC 3.2.1.4) (Béguin *et al.*, 1985; Fukumori *et al.*, 1986), maltase (α-D-glucosidase; EC 3.2.1.20) (Hong & Marmur, 1986) and isomaltase (oligo-1,6-glucosidase; EC 3.2.1.10) (Hunziker *et al.*, 1986). However, no significant homologies were found (data not shown).

A molecular model and catalytic residues of Taka-amylase A from *Aspergillus oryzae* were reported by Matsuura *et al.* (1984). Using the alignment of amino acid residues in homologous regions (Fig. 3), this model was applied for other α-amylases, the neopullulanase, isoamylase, pullulanase and cyclodextrin glucanotransferases (Fig. 4). Glu230 and Asp297 were proposed as active sites of Taka-amylase, and His122, Asp206, His210 and His296 were proposed as substrate-binding sites (Matsuura *et al.*, 1984). These active and substrate-binding sites were surprisingly conserved in other α-amylases, the neopullulanase isoamylase, pullulanase and cyclodextrin glucanotransferases (Figs 3 and 4). However, one substrate-binding-site residue, corresponding to His210 of Taka-amylase, was substituted by Glu, Val or Tyr in the neopullulanase, isoamylase or pullulanase, respectively. It is interesting that only these three enzymes can hydrolyse α-(1→6)-glucosidic linkages (Table 1).

Plant *et al.* (1987) suggested the presence of a carboxyl group with p*Ka* 4·3 at the active site of *Thermoanaerobium* pullulanase. It was considered most likely that the p*Ka* value of 4·3 represented the ionization of a Glu or Asp carboxyl group, and this view was confirmed by chemical modification with a carboxyl-group-specific reagent, 1-(3-dimethylaminopropyl)-3-ethylcarbodi-imide (Plant *et al.*, 1987). This finding is consistent with the hypothesis that Glu and/or Asp might be active-site residues in all the amylolytic enzymes shown in Figs 3 and 4.

To further investigate the role of the amino acid residues for catalytic activity and action pattern of the neopullulanase, and also to improve the characterization of the enzyme at the molecular level, amino acid substitution in the probable active site by site-directed mutagenesis is now in progress.

REFERENCES

ALLEN, G. (1981). Sequencing of proteins and peptides. In *Laboratory Technique in Biochemistry and Molecular Biology*, pp. 161–236. Edited by T. S. Work & R. H. Burdon. Amsterdam: North-Holland Publishing Co.

AMEMURA, A., CHAKRABORTY, R., FUJITA, M., NOUMI, T. & FUTAI, M. (1988). Cloning and nucleotide sequence of the isoamylase gene from *Pseudomonas amyloderamosa* SB-15. *Journal of Biological Chemistry* **263**, 9271–9275.

BÉGUIN, P., CORNET, P., & AUBERT, J.-P. (1985). Sequence of a cellulase gene of the thermophilic bacterium *Clostridium thermocellum*. *Journal of Bacteriology* **162**, 102–105.

BINDER, F., HUBER, O. & BÖCK, A. (1986). Cyclodextrin-glycosyltransferase from *Klebsiella pneumoniae* M5a1: cloning, nucleotide sequence and expression. *Gene* **47**, 269–277.

FUKUMORI, F., SASHIHARA, N., KUDO, T. & HORIKOSHI, K. (1986). Nucleotide sequences of two cellulase genes from alkalophilic *Bacillus* sp. strain N-4 and their strong homology. *Journal of Bacteriology* **168**, 479–485.

HATTORI, M. & SAKAKI, Y. (1986). Dideoxy sequencing method using denatured plasmid templates. *Analytical Biochemistry* **152**, 232–238.

HONG, S. H. & MARMUR, J. (1986). Primary structure of the maltase gene of the *MAL6* locus of *Saccharomyces carlsbergensis*. *Gene* **41**, 75–84.

HUNZIKER, W., SPIESS, M., SEMENZA, G. & LODISH,

H. F. (1986). The sucrase–isomaltase complex: primary structure, membrane-orientation, and evolution of a stalked, intrinsic brush border protein. *Cell* **46**, 227–234.

IMANAKA, T. & KURIKI, T. (1989). Pattern of action of *Bacillus stearothermophilus* neopullanase on pullulan. *Journal of Bacteriology* **171**, 369–374.

IMANAKA, T., TANAKA, T., TSUNEKAWA, H. & AIBA, S. (1981). Cloning of the genes for penicillinase, *penP* and *penI*, of *Bacillus licheniformis* in some vector plasmids and their expression in *Escherichia coli*, *Bacillus subtilis*, and *Bacillus licheniformis*. *Journal of Bacteriology* **147**, 776–786.

IMANAKA, T., FUJII, M., ARAMORI, I. & AIBA, S. (1982). Transformation of *Bacillus stearothermophilus* with plasmid DNA and characterization of shuttle vector plasmids between *Bacillus stearothermophilus* and *Bacillus subtilis*. *Journal of Bacteriology* **149**, 824–830.

KATSURAGI, N., TAKIZAWA, N. & MUROOKA, Y. (1987). Entire nucleotide sequence of the pullulanase gene of *Klebsiella aerogenes* W70. *Journal of Bacteriology* **169**, 2301–2306.

KIMURA, K., KATAOKA, S., ISHII, Y., TAKANO, T. & YAMANE, K. (1987). Nucleotide sequence of the β-cyclodextrin glucanotransferase gene of alkalophilic *Bacillus* sp. strain 1011 and similarity of its amino acid sequence to those of α-amylases. *Journal of Bacteriology* **169**, 4399–4402.

KOZAK, M. (1983). Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. *Microbiological Reviews* **47**, 1–45.

KURIKI, T., OKADA, S. & IMANAKA, T. (1988*a*). New type of pullulanase from *Bacillus stearothermophilus* and molecular cloning and expression of the gene in *Bacillus subtilis*. *Journal of Bacteriology* **170**, 1554–1559.

KURIKI, T., PARK, J.-H., OKADA, S. & IMANAKA, T. (1988*b*). Purification and characterization of thermostable pullulanase from *Bacillus stearothermophilus* and molecular cloning and expression of the gene in *Bacillus subtilis*. *Applied and Environmental Microbiology* **54**, 2881–2883.

MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (editors) (1982). Gel electrophoresis. In *Molecular Cloning: a Laboratory Manual*, pp. 149–185. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

MATSUMURA, M., KATAKURA, Y., IMANAKA, T. & AIBA, S. (1984). Enzymatic and nucleotide sequence studies of a kanamycin-inactivating enzyme encoded by a plasmid from thermophilic bacilli in comparison with that encoded by plasmid pUB110. *Journal of Bacteriology* **160**, 413–420.

MATSUURA, Y., KUSUNOKI, M., HARADA, W. & KAKUDO, M. (1984). Structure and possible catalytic residues of Taka-amylase A. *Journal of Biochemistry* **95**, 697–702.

MCPHERSON, M. J. & CHARALAMBOUS, B. M. (1988). Functional analysis of the starch debranching enzyme pullulanase. *Biochemical Society Transactions* **16**, 723–724.

MORAN, C. P., JR, LANG, N., LEGRICE, S. F. J., LEE, G., STEPHENS, M., SONENSHEIN, A. L., PERO, J. & LOSICK, R. (1982). Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. *Molecular and General Genetics* **186**, 339–346.

NAKAJIMA, R., IMANAKA, T. & AIBA, S. (1985). Nucleotide sequence of the *Bacillus stearothermophilus* α-amylase gene. *Journal of Bacteriology* **163**, 401–406.

NAKAJIMA, R., IMANAKA, T. & AIBA, S. (1986). Comparison of amino acid sequences of eleven different α-amylases. *Applied Microbiology and Biotechnology* **23**, 355–360.

PERRY, L. J. & WETZEL, R. (1984). Disulfide bond engineered into T4 lysozyme: stabilization of the protein toward thermal inactivation. *Science* **226**, 555–557.

PLANT, A. R., CLEMENS, R. M., MORGAN, H. W. & DANIEL, R. M. (1987). Active-site and substrate-specificity of *Thermoanaerobium* Tok6-B1 pullulanase. *Biochemical Journal* **246**, 537–541.

RHODES, C., STRASSER, J. & FRIEDBERG, F. (1987). Sequence of an active fragment of *B. polymyxa* beta amylase. *Nucleic Acids Research* **15**, 3934.

ROSENBERG, M. & COURT, D. (1979). Regulatory sequences involved in the promotion and termination of RNA transcription. *Annual Review of Genetics* **13**, 319–353.

SAKAI, S., KUBOTA, M., YAMAMOTO, K., NAKADA, T., TORIGOE, K., ANDO, O. & SUGIMOTO, T. (1987). Cloning of cyclodextrin glucanotransferase genes from *B. stearothermophilus* and *B. macerans*. *Journal of the Japanese Society of Starch Science* **34**, 140–147.

SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5463–5467.

TAKAGI, M., IMANAKA, T. & AIBA, S. (1985). Nucleotide sequence and promoter region for the neutral protease gene from *Bacillus stearothermophilus*. *Journal of Bacteriology* **163**, 824–831.

TINOCO, I., JR, BORER, P. N., DENGLER, B., LEVINE, M. D., UHLENBECK, O. C., CROTHERS, D. M. & GRALLA, J. (1973). Improved estimation of secondary structure in ribonucleic acids. *Nature New Biology* **246**, 40–41.

VIEIRA, J. & MESSING, J. (1987). Production of single-stranded plasmid DNA. *Methods in Enzymology* **153**, 3–11.

YAMASHITA, I., SUZUKI, K. & FUKUI, S. (1985). Nucleotide sequence of the extracellular glucoamylase gene *STA1* in the yeast *Saccharomyces diastaticus*. *Journal of Bacteriology* **161**, 567–573.